GENETIC BASED INTRUSION DETECTION SYSTEM IN ADVANCED METERING INFRASTRUCTURE

Ali Saeedi
Mashhad Electric Energy Distribution Company (MEEDC), Iran

Mohammad Hossein Yaghmaee
Mashhad Electric Energy Distribution Company (MEEDC), Iran
yaghmaee@ieee.org

Niki Sagharidooz
Mashhad Electric Energy Distribution Company (MEEDC), Iran
n.sagharidooz@imamreza.ac.ir

ABSTRACT

Advanced metering infrastructure (AMI) is a fundamental element of the smart grid. AMI includes varied networks, systems and communications media that permits attackers interfere with communications and compromises utility devices or steal customers' private data. Therefore, the security of AMI is an important topic in the implementation of smart grid. One of the critical needs for security is monitoring the traffic to detect intrusions and make reports. Most of intrusion detection systems (IDS) in AMI are signature-based which neither detect unknown attacks, nor have data for false positive rate (wrong alarm). In this paper, first, data mining classification algorithm is used to detect known attacks and second, rule-based IDS is applied in genetic algorithm to detect unknown attacks. The proposed IDS architecture includes intrusion detection mechanisms for detecting compromised smart meters, data collectors and Head-ends (HAN, NAN and WAN networks). The results illustrate noticeably higher detection and exceptionally low false positive rate.

INTRODUCTION

The digital technology that provides two-way communication between power producers and customers along the transmission line is smart grid. It is a unique opportunity to enter a new level of reliability, availability and efficiency that will help the economy, health and the environment. Advanced Metering Infrastructure (AMI) is designed with the aim of providing a comprehensive structure to control and monitor power distribution part which consumers can observe and control their power consumption in near real time. The security requirements of AMI are confidentiality, integrity, availability and Non-Repudiation [1]. In order to satisfy the requirements, AMI uses physical protection, meter authentication, encryption of communications, firewalls and Intrusion Detection Systems (IDS) [2].

Our goal is designing an intrusion detection system, by applying genetic algorithm (GA) to detect known and unknown intrusions efficiently. Proposed method includes two phases; in training phase, best classification rule set are produced by genetic algorithm and data mining which are stored in a rules database; in detection phase, produced rules are used to classify traffic data of AMI. Result of classification is 99.94% correctly classified and 0.05% is wrongly classified.

The rest of this paper is explained about related works to IDS in AMI then Background information about IDS, KDDcup99 dataset and genetic algorithm are presented. After that, proposed IDS in AMI is described and next simulation and evaluation results are illustrated and finally conclusion of proposed method is brought.

RELATED WORKS

Intrusion detection system is a monitoring technique that warns about unauthorized entry into the network. Robin Berthier et al. [3] discussed about detection technologies which are signature-based, anomaly-based and specification-based detection. Because signature-based detection uses blacklist, it cannot detect unknown attacks and needs frequently updates. Authors of paper proposed specification-based detection for AMI; although it has high accuracy, its implementation is really costly and less scalable than anomaly detection [3, 4]. Due to problems with other detection technologies, in this paper anomaly-based detection by data mining method is used.

David Grochoki et al. [5] presented a survey of threats and attack techniques which occurs in AMI. Also they investigated about IDS deployment in AMI that they suggested hybrid sensing infrastructure approach to provide a wide coverage in monitoring attacks. This approach uses both centralized IDS and embedded meter sensors.

Mustafa Amir Faisal et al. [6] proposed IDS architecture for AMI which consists of three IDSs deployed in smart meters, data concentrators, and central system. They used data stream mining for detecting anomaly in AMI.

Robert Mitchell et al. [7] introduced a behavior-rule based intrusion detection for protection of head-ends, distributed access points, subscriber energy meters of smart grid which supports secure applications for subscribers. The proposed IDS have high detection rate (near 100%) and high false positive rate (e.g., less than 10%). According to Mitchell, IDS techniques for AMI are yet in their beginnings with just a little work reported in the literature and most of them don't have false positive rate and ROC plot.

Yichi Zhang et al. described a method for distributed IDS in smart grid by deploying smart module in multi-layer of smart grid. Each of these modules analyses data related to their layer and communicates with other layer**.** For classify malicious data, Support Vector Machine

(SVM) and Artificial Immune System (AIS) are used.

INTRUSION DETECTION SYSTEM

IDS is a computer system which attempts to identify actions that try to compromise the integrity, confidentiality or availability of a computer resource. When an intrusion is detected, the system reports by an alarm to an operator [9].

Networking Attacks

This part is an overview of the four major classes of networking attacks [10].

Denial of Service (DoS)

A DoS attack is an attack in which the hacker can make the system unusable or slow down its speed by imposing an additional burden on system resources, so normal users may not be able to work with the server. E.g. apache, smurf, neptune, ping of death, back, mail bomb, UDP storm etc. which are all DoS attacks.

Remote to User Attacks (R2L)

Thorough the internet, Attacker sends packets to a machine which s/he does not have access to. Then s/he exposes vulnerabilities of machines in order to exploit privilege. Consequently, attacker would be like a local user. Examples of R2L are xlock, guest, xnsnoop, phf, sendmail dictionary etc.

User to Root Attacks (U2R)

Attacker is normal user which abuse her/his privilege e.g. perl, xterm. Detection of these attacks is very hard.

Probing

Hacker scans a machine or a network device with the purpose of determining its weaknesses or vulnerabilities that may help to compromise the system. E.g. saint, portsweep, mscan, nmap etc. which are all probing attacks.

KDDcup 1999

KDDcup99 is a dataset which includes a wide variety of intrusions simulated in a military network environment. In this paper KDDcup99 is used to differentiate bad connection (attack) from good connection (normal).

In table I, the attack names of KDDcup99 are brought.

Table I. Attack Names of KDDcup99

Attack classes	attack names
DOS	back, land, neptune, pod, smurf, teardrop
U2R	buffer_overflow, perl, loadmodule, rootkit
R2L	ftp_write, guess_passwd, imap, multihop, phf, spy, warezclient, warezmaster
Probing	ipsweep, nmap, portsweep, satan

GENETIC ALGORITHM

Genetic algorithms are optimizations algorithm that mimics both natural selection and natural genetics. In the trend of GA, there are three operations that are, selection, crossover and mutation. First, initial random individuals (chromosomes) are generated and are given to a fitness function to evaluate its goodness. Second, individuals with high fitness value are selected in the selection phase. Third, in crossover phase, each two selected individuals are divided from a single point and then swap one part of their digits to each other. Fourth, a single bit of children from the previous step is changed randomly that is mutation phase and finally new population is created. This procedure of selection, crossover and mutation is repeated until it is reached to maximum generation number [12].

GA in IDS

Payel Gupta et al. [13] presented a survey about problems of methods in Intrusion Detection Systems such as clustering technique (ex. K-Means), SVM, Naïve Bayes classifier, etc.; they also used a linear classification to implement anomaly-based IDS, but because of high false positive, they applied genetic algorithm that is based on if-then rules. For the advantage of GA, they said, “GAs is robust, inherently parallel, adaptable and suitable for dealing with the classification of rare classes.” (123).

Mohammad Sazzadul Hoque et al. [14] introduced an IDS that included pre-calculation and detection parts. In pre-calculation step, by using network audit data (train data), a set of chromosomes produced. In detection step, network audit data (test data) and chromosomes considered as input of GA algorithm so the type of the test data is predicted.

Ojugo et al. [15] proposed a genetic algorithm method to create set of rules from audit network data and they added support-confidence framework to evaluate the rules in fitness function.

PROPOSED IDS FOR AMI

In this section, proposed anomaly-based Intrusion detection system is described which consists of two phases: training and detection. (See Figure 1)

Training Phase

In this phase, first, data mining classification technique is used to classify dataset and rules for known intrusions are created by using C4.5 algorithm. Second, genetic algorithm is applied to create the best classification rule set for detecting unknown attacks. In order to produce new rules, GA uses dataset for initial population and then with selection, crossover and mutation operations, new population is created and this process is continued until the termination condition is reached. Third both of rules are stored in a rules database.

Detection Phase

In each network, there is an IDS sensor which analyses the traffic then Intrusions are detected and normal traffic is sent to next network. Produced rules from previous step are used to differentiate attacks from normal data.

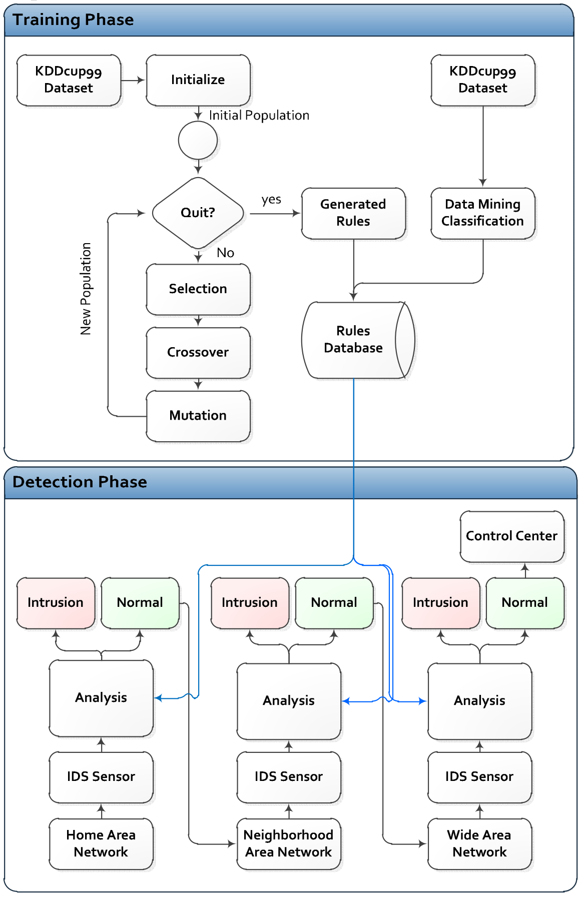

Figure 1.Intrusion detection process in AMI

Genetic algorithm in proposed IDS

In this section the proposed GA based IDS system is described.

Features Selection

There are 41 features in KDDcup99 dataset. To reduce the computation, Information Gain Ranking Filter algorithm is applied for attribute evaluator and ranker as search method and then features with low rank are eliminated. It is simulated in WEKA and results of 12 best features are as follows:

src_bytes, count, service, srv_count, dst_host_same_src_port_rate, protocol_type, dst_host_srv_count, dst_host_diff_srv_rate, flag, dst_host_same_srv_rate, diff_srv_rate, same_srv_rate,

Encoding

Selected features are determined as genes of a chromosome and each of the chromosomes is a rule. Rules are usually presented in the following form:

If {condition} then {act}

Ex: ***If*** *src_bytes > 22 AND*
flag = RSTR AND
dst_host_diff_srv_rate <= 0 AND
service = http ***then*** *back (attack name)*

12 selected features are defined as the condition part and attack name feature as the act part and features are encoded based on their length.

Fitness Function

In the training phase of GA, Chromosomes are evaluated by fitness function to determine its goodness. If the chromosome properly classifies an attack, its fitness value should be greater. The best rule is a chromosome with highest fitness value. Support-confidence framework is used to create the fitness function. If we have the rule:

If A then B, then
Support = |A and B| / N
Confidence = |A and B| / |A|
*Fitness = w1 * support + w2 * confidence*
N = Number of connections in dataset
|A| = Number of connections matching condition A.
|A and B| = Number of connections matching rule, if A and B
w1, w2 = Weights to balance the two terms.

Selection

Tournament selection is used in which chromosomes are randomly chosen from current population and several tournaments are run among chromosomes.

Crossover

Two-point crossover algorithm is set for crossover step that chooses two random cross section points from the chromosome and swaps the anything between the parents.

Mutation

Depending on the probability of mutation rate, some gens of chromosomes are randomly changed.

SIMULATION AND EVALUATION RESULS

To evaluate the performance of proposed IDS, the WEKA data mining tool is used to simulate and evaluate classification algorithms. In WEKA, a clone of C4.5 algorithm is used, which is named J48. 66% of KDDcup99 dataset is given to classifier as train data and 34% as test data. Result of J48 classification is 99.94% correctly classified and 0.05% is wrongly classified.

Evaluation Criteria

Detail results of classification by J48 algorithm are brought in table II.

True Positives (**TP**): is the number of normal connections classified as normal.

False Positives (**FP**): is the number of normal connections classified as attacks.

True Negatives (**TN**): is the number of attack connections classified as attacks.

False Negatives (**FN**): is the number of attack connections classified as normal.

$$Recall = \frac{TruePositive}{FalseNegetive + TruePositive}$$

$$precision = \frac{TruePositive}{TruePositive + FalsePositive}$$

Table II. Results of classification by J48 algorithm

	True Positive	False Positive	Precision	Recall	Class
	0.997	0	0.995	0.997	Back
	0.615	0	0.8	0.615	buffer_overflow
	0	0	0	0	ftp_write
	0.952	0	1	0.952	guess_passwd
	0.8	0	1	0.8	Imap
	0.988	0	0.988	0.988	Ipsweep
	1	0	0.8	1	Land
	0	0	0	0	loadmodule
	0	0	0	0	multihop
	1	0	1	1	neptune
	0.926	0	0.949	0.926	nmap
	0.999	0	0.998	0.999	normal
	0	0	0	0	perl
	0.5	0	1	0.5	phf
	1	0	0.987	1	pod
	0.979	0	0.987	0.979	portsweep
	0	0	0	0	rootkit
	0.986	0	0.989	0.986	satan
	1	0	1	1	smurf
	0	0	0	0	spy
	0.997	0	1	0.997	teardrop
	0.953	0	0.991	0.953	warezclient
	0.75	0	0.75	0.75	warezmaster
Avg.	0.999	0	0.999	0.999	

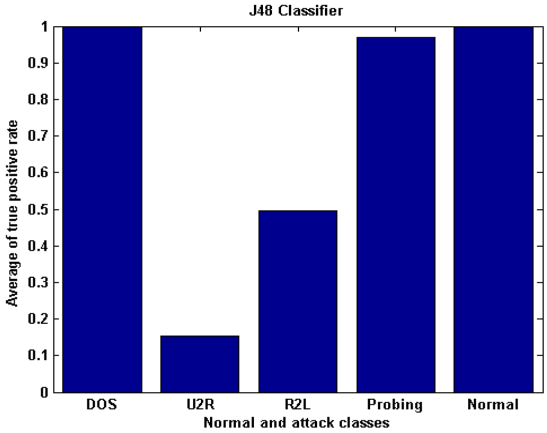

Figure 2. Detected attacks by J48 classifier

Comparaison of Classifiers

In this section, five classifier algorithms are selected which are : J48, NaiveBayes, OneR, Part and ZeroR. Because of limitations in AMI, these algorithms are compared to the percentage of correctly classification (Figure 3), the size of each model (Figure 4) and time taken to build a model of rules (Figure 5). Part and J48 classifiers have high correctly detection rate, but size of rules model and time taken for building Part model has the highest value in two last figure. OneR classifier also has a high detection rate of correct instances and very low value in time and size of train model, but the false positive rate of OneR algorithm is higher than J48 so J48 is suggested to use. According to the simulations, NaiveBayes and ZeroR classifiers are not good enough.

Each algorithm is run with different percentage of training data and remains percentage is for testing data. The results of each algorithm are so close to each other.

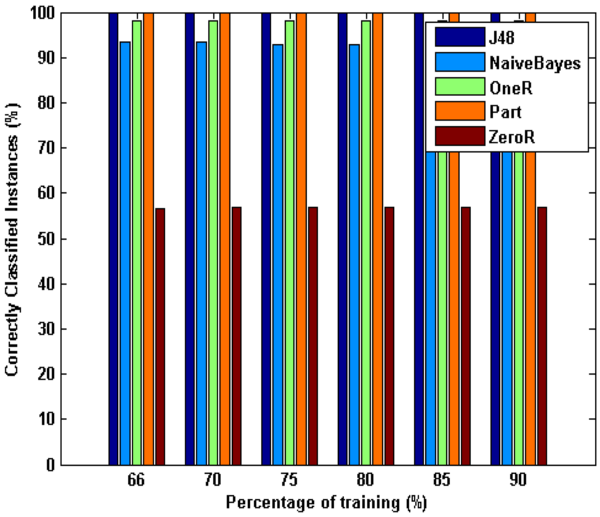

Figure 3. Percentage of correctly classification of five algorithms in each percentage of training

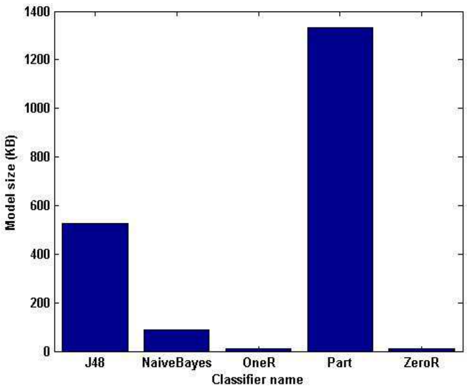

Figure 4. Classification Model size of five algorithms

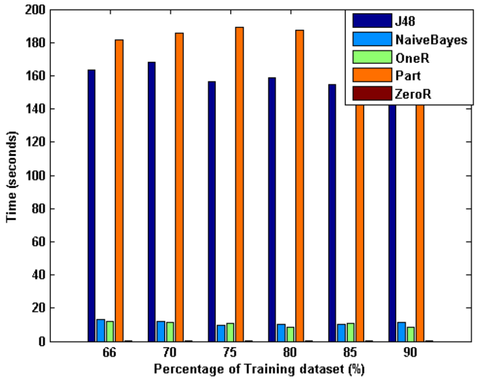

Figure 5. Time taken to build a model of five algorithms in each percentage of training

CONCLUSION

This paper presents Intrusion Detection System (IDS) based on genetic algorithm and data mining approach for advanced metering infrastructure. The proposed IDS can detect known and novel attacks by rules database in each network of AMI, then the normal data are passed through the next network (ex. HAN to NAN), at last, just normal traffics are reached to control center. To measure goodness of rules that are created in genetic algorithms, the support - confidence framework of association rule mining is used. Different classification algorithms are also compared to each other for their accurate detection, size of the model and time taken to create rules of training data. J48 classifier with applying KDDcup99 dataset has a high detection rate of DoS and Probing attacks and very low false positive rate (0.05%).

REFERENCES

[1] F.M. Cleveland, 2008, "Cyber security issues for advanced metering infrastructure" *Proceedings IEEE Power and Energy Society General Meeting-Conversion and Delivery of Electrical Energy*, 1-5.

[2] Stephen McLaughlin, Dmitry Podkuiko, Sergei Miadzvezhanka, Adam Delozier, Patrick McDaniel, 2010, "Multi-vendor penetration testing in the advanced metering infrastructure", *Proceedings of the 26th Annual Computer Security Applications Conference*, 107-116

[3] Robin Berthier, William H. Sanders, and Himanshu Khurana, 2010 "Intrusion Detection for Advanced Metering Infrastructures: Requirements and Architectural Directions", *Proceedings First IEEE International Conference on Smart Grid Communications*, 350-355.

[4] Robin Berthier and William H. Sanders, 2011, "Specification-based Intrusion Detection for Advanced Metering Infrastructures", *Proceedings IEEE 17th Pacific Rim International Symposium on on Smart Grid Communications*, 184-193.

[5] David Grochocki, Jun Ho Huh, Robin Berthier, Rakesh Bobba and William H. Sanders, 2012, "AMI Threats, Intrusion Detection Requirements and Deployment Recommendations", *Proceedings IEEE Third International Conference on Smart Grid Communications*, 395-400.

[6] Mustafa Amir Faisal, Zeyar Aung, John R. Williams and Abel Sanchez, 2012, "Securing Advanced Metering Infrastructure Using Intrusion Detection System with Data Stream Mining", *Intelligence and Security Informatics,ISI*, vol. 7299, 96-111.

[7] Robert Mitchell and Ing-Ray Chen, 2013, "Behavior-Rule Based Intrusion Detection Systems for Safety Critical Smart Grid Applications", *Smart Grid, IEEE Transactions, TGC*, vol. 4, 1254-1263.

[8] Yichi Zhang, Lingfeng Wang, Weiqing Sun, Green, R.C., Alam, M., 2011, "Distributed intrusion detection system in a multi-layer network architecture of smart grids", *IEEE Tran. Smart Grid, TGC*, vol. 2, 796-808.

[9] Florian Kerschbaum, Eugene H. Spafford, Diego Zamboni, 2002, "Using internal sensors and embedded detectors for intrusion detection", *Journal of Computer Security, JCS*, vol. 10, 23-70.

[10] A. Sung, S. Mukkamala, 2003, "Identifying important features for intrusion detection using support vector machines and neural networks", *Symposium on Applications and the Internet, SAINT*, 209-216.

[11] KDD Cup 1999 Dataset, available [online]: http://kdd.ics.uci.edu/databases/kddcup99/kddcup99.html

[12] David A Coley, 1999, *An introduction to genetic algorithms for scientists and engineers*, world scientific publishing Co. Pt. Ltd., River Edge, New Jersey, 224

[13] Payel Gupta and Subhash K. Shinde, 2011, "Genetic Algorithm Technique Used to Detect Intrusion Detection", *Proceedings of the First International Conference on Advances in Computing and Information Technology*, 122-131.

[14] Mohammad Sazzadul Hoque, Md. Abdul Mukit and Md. Abu Naser Bikas, 2012, "An Implementation of Intrusion Detection System Using Genetic Algorithm", *International Journal of Network Security & Its Applications, IJNSA*, vol. 4, 109-120.

[15] A.A. Ojugo, A.O. Eboka, O.E. Okonta, R.E Yoro (Mrs), F.O. Aghware, 2012, "Genetic Algorithm Rule-Based Intrusion Detection System (GAIDS) ", *Journal of Emerging Trends in Computing and Information Sciences, CIS*, vol. 3, 1182-1194.